

MEASURES OF QUALITY OF LIFE AMONG UNIVERSITY STUDENTS

Isabella Sulis, Nicola Tedesco¹

Dipartimento di Ricerche Economiche e Sociali, Università di Cagliari, Cagliari, Italy

Abstract *This article outlines findings from a survey addressed to measure the quality of life of university students in Cagliari. It focuses on issues related to the process of building up of a synthetic indicator of students' quality of life from responses to a set of subjective indicators all measured on ordered scale. The aim has been pursued by adopting the modeling approach of the Item Response Models which enable us to simultaneously summarize student's multiple responses in a metrical measure of the latent trait and to assess the properties of each indicator in terms of the location of its parameters on the latent trait and its capability to discriminate across students. A comparison analysis with other classical scaling procedures to summarize multiple indicators in a single statement has been carried out with the main aim to assess the potential of the Item Response Models approach in terms of capability to detect pattern of responses which signal a different intensity of the latent trait.*

Keywords: *Item response models, Students' quality of life, Mixed-effects models, Synthetic indicator*

1. INTRODUCTION

Since 2001 the Italian university system has faced a phase of overall reorganization in a new formative system based on two levels (first level- three years - and second level - two years - degree). In this phase the efforts of the researchers have been mainly addressed to measure the efficiency and the effectiveness of the academic system monitoring students' transition from the university to the job-market (Balbi and Grassia, 2007; Porcu and Tedesco, 2007), the regularity in their curricula and drop-outs (Biggeri and Bini, 2001; Bini and Bertaccini, 2007; Chiandotto and Bacci, 2007; Porcu and Puggioni, 2003), and the quality of university courses (Bernardi et al., 2004; Capursi and Porcu, 2001; Rampichini et al., 2004; Sulis, 2007). Fewer attention has been paid to the analysis of the attitude of the institutions towards students' requests and the actions addressed in favour of them. An aspect which has been marginally analyzed is the quality of life of students during their permanence in the university system and its influence on their academic

1 Isabella Sulis, email: isulis@unica.it

performances.

In the Italian framework the concept of quality of life of university students has been analyzed more in terms of adequacy of structures, facilities and the effect of social environment on students' well-being (Aureli and Grimaccia, 1999; Maggino and Schifini, 1999) than in terms of students' habits of life during the university studies. Researchers agree in defining students' quality of life at the university as a latent variable that can be measured through the use of *objective* and *subjective* indicators. The latter are addressed to reveal psychological and individual aspects (Huebner et al., 2005; Larsen et al., 1985). The measurement process requires firstly the latent variable to be decomposed in terms of dimensions and next to choose, according to some rational criteria, a set of indicator variables for each dimension (Cox et al., 1992; Fayers and Hand, 2002). The definition of such aspects is not straightforward and the choice of the components is influenced by many factors: availability of the information; the geographical context in which the university is located; socio-cultural factors; economic conditions of the area etc. Furthermore the process of measurement implies a high level of arbitrariness in the phase of definition of many components (indicator variables, transformation functions, margin functions and weighing system) of the phenomena that are usually left to single researchers' choices.

On the other hand, the 'objective approach' moves from the identification of a set of factual variables that indicate the standard of life (level of income, the ownership of a house or of particular goods) and defines the procedures of synthesis of the single components. In the 'subjective approach' the stress is on the level of well-being as perceived by single individual (independently from his/her objective standard of life). Mixed strategies use both 'subjective' and 'objective' indicators since they are more apt to highlight the phenomenon under different points of views (Aureli and Grimaccia, 1999; Maggino and Schifini, 1999; Shulz, 1999). An interesting aspect of the mixed strategies arises from the inspection of the correlation between the two components (subjective and objective).

This work moves from a *subjective* prospective. We consider objective indicators (as for instance the *social-economic conditions*) confounding factors which could have influenced the ratings observed for the set of subjective indicator variables. Specifically, students' quality of life during the university studies has been monitored by means of subjective indicator variables. We advance the hypothesis that the quality of life of a student is mainly determined by his/her style of life and by his/her level of integration in the academic and city environment. We suppose students who have the highest level of quality of life take advantage of all the services of the university, are perfectly integrated in the city and enjoy their students'

status by taking part to many external activities.

A questionnaire addressed to know students' habits of life has been used as tool for operationalizing the underlying attribute. Broadly speaking the indicator items selected provide information on students' habits in their daily life at the university.

A modeling approach for the assessment of 'students' quality of life' has been built up and tested on a sample of students' enrolled to three faculties of the University of Cagliari: Economics, Law and Political Sciences. These three faculties have been selected since their students are supposed to be enough homogenous in respect of several characteristics: similar formative curricula and similar location of faculties (the same area of the city and contiguous buildings and several common spaces). The research mainly focuses on two aspects:

1. to determine the level of students' quality of life at the university moving from a bunch of *subjective* indicators;
2. to make a comparison study between several scaling methods.

Section 2 provides information on the data and variables adopted in our analysis. Section 3 presents the methodological approach and discusses the main results of a comparative analysis. Section 4 outlines the differences with other classical scaling methods for metrical data. Section 5 presents our conclusions.

2. THE DATA AND THE INDICATORS

This survey considers as reference population 13893 students enrolled at the faculties of Economics, Law and Political Sciences in 2001/2002 a.y. (36.9% of students at the University of Cagliari). A quota sample has been carried out using a two stage sampling procedure in respect of the variables 'faculty' (Law, Political Sciences, Economics) and gender (M, F). Finally, within each of the six strata a stratified non proportional sample (with a constant number of students) has been selected according to the 'residential status'. Using the information on the distance between students' accommodation and the location of the University three groups of students have been detected: 'resident', 'non resident', 'commuter'. A final sample size of 375 units has been obtained. The rate of sampling is equal to 2,7% of the overall population: 43.2% belongs to Law, 35.2% to Economics and 21,6 to Political Sciences. The distribution of students conditionally upon their 'age' and 'academic status' ('first year student', 'regular student', 'no regular student') appears to be enough balanced between female and male. It is worth pointing out that 'student status' and 'age' have a specular distribution since the former

Table 1: Indicator items selected to measure students' quality of life: distributions of students' responses

item and meaning	%		
	never	sometimes	often
<i>cus</i> - use of university sporting center	69.60	19.47	10.93
<i>theater</i> - attendance at theater	77.07	19.73	3.20
<i>canteen</i> - use of the university refectory	66.67	21.33	12.00
<i>sport</i> - practising a sporting activity	45.33	25.07	29.60
<i>cultural</i> - attendance at cultural events	60.27	30.93	8.80
<i>reading</i> - reading non academic books	33.87	32.80	33.33
<i>work</i> - full time, part time, no job	44.80	35.20	20.00
<i>disco</i> - attendance at disco	46.93	37.07	16.00
<i>lectures</i> - attendance at lectures	11.20	38.13	50.67
<i>bar</i> - attendance at bar	18.13	40.80	41.07
<i>meeting</i> - meeting with lecturers	50.93	42.67	6.40
<i>clubbing</i> - attendance at clubs	22.13	44.00	33.87
<i>library</i> - use of university library	15.73	51.47	32.80
<i>cinema</i> - attendance at cinema	25.33	58.13	16.53

can be considered a proxy of the latter. The questionnaire used in the survey is structured in sections which provide information on several aspects: students' personal details and the social-economic status of his/her family; students' attitude to use the university facilities and students' style of life in Cagliari. Some of these indicators are applicable just to 'commuter' and 'non resident' students. Other items are addressed to reveal directly or indirectly students' economic conditions (type of accommodation; how much he/ she spends for accommodation, food, etc; type of transport frequently used; if he/she owns a vehicle and type of vehicle; etc). The main part of the questionnaire is composed of questions measured on categorical ordered scale addressed to know how often students are involved in specific activities. This set of items are classified as subjective indicators since they move from individual responses in order to define students' level of quality of life at the university. According to the hypothesis followed in this research, the higher the number of aspects on which they take part, the more they are involved in the social and academic life.

The meaning of the set of 14 items selected to know how often students attend or use services is reported in Table 1.

Excluding 'work' all of them have a positive direction. In our hypothesis to

Table 2: Rate of positive answers by adopting two different dichotomization rules

item	%			
	often No	sometimes-never Yes	often-sometimes No	never Yes
theater	96.80	3.20	8.80	91.20
meeting	93.60	6.40	6.40	93.60
cultural	91.20	8.80	41.07	58.93
cus	89.07	10.93	10.93	89.07
canteen	88.00	12.00	12.00	88.00
disco	84.00	16.00	33.33	66.67
cinema	83.47	16.53	3.20	96.80
no work	80.00	20.00	29.60	70.40
sport	70.40	29.60	16.53	83.47
library	67.20	32.80	32.80	67.20
reading	66.67	33.33	33.87	66.13
clubbing	66.13	33.87	86.67	13.33
bar	58.93	41.07	16.00	84.00
lectures	49.33	50.67	50.67	49.33

have a ‘part-time’ or ‘full-time’ job means to have less time to devote to other academic and non academic activities; so we changed the direction of this item. At the same time we use as control variables all items concerning students’ *social-economic conditions* and *academic curricula*. Table 1 exhibits the observed rates of responses for each category of the 14 indicators. It is interesting to stress the unexpected high rate of students who use the category ‘never’ for the items concerning the attendance at ‘cultural events’ (60.3%) and ‘theater’ (77.1%), ‘meeting’ (50.9%), ‘sport’ (45.3%) and ‘reading’ (33.9%). As a first attempt to explore students’ response pattern we dichotomized the indicators collapsing the grade of the scale into two categories ‘yes’ and ‘no’ according to two different rules: merging ‘sometimes with yes’ and ‘sometimes with no’ (see Table 2).

Table 2 points out the arbitrariness in the responses arisen postulating questions as binary. The value of the *Spearman* correlation coefficient ($\rho = -0.49$) between the two rankings highlights the meaningless of using methods for binary data: ‘theater’ switches from the first to the twelfth rank, ‘meeting’ from the second to the thirteenth and so forth. In the following the specific modeling approach of the *Item Response Models* (Baker and Kim, 2004; Masters, 1982; Muraki, 1992; Samejima, 1969) for polytomous ordered items will be adopted

to estimate measures on a metrical scale of the latent trait ‘students’ quality of life’. The individual metrical measures obtained by adopting the IRM approach will be compared with two indicators built up by using classical scaling methods which assign equal weight to the items involved in the analysis.

3. A COMPARISON ACROSS MODELS

In *Item Response Theory* the probability for subject p ($p = 1, \dots, n$) to score category k of item i ($i = 1, \dots, I$) is function of the *item-category parameters*² and *person parameter*; a further parameter, known as *discrimination parameter*, helps to differentiate across items with different discrimination power. *Person parameters* provide an individual measurement of students’ standard of life. *Item-category parameters* locate the ordered categories in the continuum. *Discrimination parameters* provide a weighing scheme to summarize the indicator items in a synthetic indicator. One of the more widely adopted Item Response Model (IRM) to handle ordered polytomous items is the Graded Response Model (GRM) (Samejima, 1969)

$$P(Y_{ip} \leq k|\theta) = \frac{\exp(\tau_{ik} - \alpha_i \theta_p)}{1 + \exp(\tau_{ik} - \alpha_i \theta_p)}, \quad j = 1, \dots, J-1. \quad (1)$$

This model adopts $K-1$ cumulative *logits* to express the ratio between the probabilities to score for item i category k (for $k = 1, \dots, K$) or lower on the probability to score higher categories. The model enables *threshold parameters* τ_{ik} to differ across items. For each category k , values higher than the thresholds imply greater probabilities of responding in categories lower rather than greater than k . For each category a ranking of the items can be advanced sorting them according the value of the threshold parameter. Threshold parameters are also known as cut-points on the logistic scale that map the range of probability (0-1) onto $(-\infty, +\infty)$. In equation 1 factor loadings α_i are constrained to be constant across categories. Person parameters θ_p are specified to be random effects which vary among subjects following a $\theta_p \sim \mathcal{N}(0, \sigma_\theta^2)$ distribution (Bartholomew, 1998; De Boeck and Wilson, 2004). Model 1 is considered the random-effects version of the Proportional Odds Model (Agresti, 2002).

The factor loading α_i describes the effect of the person parameter (which measures student’s quality of life) on the cumulative probability of responding

² For polytomous items, we call *item parameters* also the *threshold parameters* which characterize the cut-point of each category (*category parameters*).

up of a category. If *discrimination parameters* are specified constant across the items, e.g. $\alpha_i = 1$ for $i = 1, \dots, I$, all questions discriminate in the same way across individuals with different *person parameters*. The negative sign on the discrimination parameter α_i indicates that as student's quality of university life increases, the higher is the likelihood for the response to fall at the high end of the scale. The effect of the person and item parameter is additive: for any item i , the higher the value of an individual on the latent trait (θ_p), the greater the probability to score higher categories.

A more parsimonious model for ordered variables is the *Proportional Odds Models* with *threshold parameters* constant across items and an item parameter β_i which shifts the cut points towards the low end of the scale

$$P(Y_{ip} \leq k | \theta) = \frac{\exp(\tau_k - \beta_i - \alpha_i \theta_p)}{1 + \exp(\tau_k - \beta_i - \alpha_i \theta_p)}, \quad j = 1, \dots, J - 1. \quad (2)$$

This is also known as the Rating Scale version (Mouraki, 1990) of the GRM. Higher values of β_i implies larger probabilities to score positive categories.

A further IRM to handle with ordinal manifest items which is characterized by the same parameters of the GRM is the Generalized Partial Credit Model (GPCM) (Muraki, 1992). It uses the adjacent-category logits link to model the item response probabilities for each item

$$P(Y_{ip} = k | \theta) = \frac{\exp \sum_{k=0}^K (\tau_{ik} + \alpha_i \theta_p)}{\sum_{r=0}^{m_i} \exp \sum_{k=0}^r (\tau_{ik} + \alpha_i \theta_p)}, \quad j = 1, \dots, J - 1; \quad (3)$$

where m_i is the number of categories of item i . When the discrimination parameters are a constant across the items the model is known as Partial Credit Model (PCM) (Masters, 1982). With adjacent-categories link function the discrimination parameters of the model describe the effect of the person parameter on the probability to answer category k rather than $k - 1$, as well the threshold parameters quantify the relative difficulty of a category compared with other categories within an item. GRM and GPCM provide similar fit in modeling ordered categorical items. A first comparison between the two models led to prefer the GRM link function, as Table 3 shows.

However, as highlighted by Agresti (Agresti et al., 2000), the choice between the two logit functions (cumulative versus adjacent-categories) is more related to whether it is more suitable to refer effects to groupings of categories using the

Table 3: Cumulative versus adjacent-categories link function

M	n	α_i	n° param.	log L	AIC	BIC
<i>GRM</i>	375	$\neq 1$	42	-4648.76	9381.52	9546.45
<i>PCM</i>	375	$=$	29	-4853.92	9765.84	9879.73
<i>GPCM</i>	375	$\neq 1$	42	-4654.11	9392.23	9557.16

Table 4: Comparisons between Graded Response Models with different characteristics in terms of goodness of fit

Model	units	$Var(\theta)$	τ_{ik}^*	α_i	n° param.	AIC	BIC
M_1	375	.62 (.07)	$=$	$= 1$	16	9880.9	9943.76
M_2	375	1.34 (.34)	$=$	$\neq 1$	29	9652.4	9766.33
M_3	375	.65(.07)	\neq	$= 1$	29	9733.0	9846.88
M_4	375	.42 (.15)	\neq	$\neq 1$	42	9381.5	9557.16

* Threshold parameters are constant across items $=$ or differ among them \neq

entire scale or to single categories; the cumulative link has the advantage that enables to refer the effects to an underlying latent variable (Agresti et al., 2000).

Bayes' theorem (Skrondal and Rabe-Hesketh, 2004) is used to get the posterior distribution of θ given the vector of observed responses \mathbf{y}

$$f(\theta|\mathbf{y}) \propto f(\theta)f(\mathbf{y}|\theta).$$

Moving from the posterior distribution, the individual parameter may be predicted on the scale of θ using as measures of location and accuracy of θ_p the $E(\theta|\mathbf{y})$ and its $Var(\theta|\mathbf{y})$.

In the following we focus on the characteristics of different GRM. Specifically, four models are set up and results compared in terms of goodness of fit, weighting schemes, ranking of the items according to their 'easiness' and interpretability. The characteristics of the four models are depicted in Table 4. The simplest model (M_1) loads all items on the latent trait with the same weight and thresholds τ_k are specified to be constant across the items; M_2 differs from M_1 since it allows factor loadings to vary and fixes the factor loading of item 'reading' equal to 1; model M_3 has thresholds which vary across items whereas factor loadings are fixed equal to 1; the most complex models M_4 leaves free both threshold parameters and factor loadings. M_4 shows the best goodness-of-fit measures (AIC and BIC), followed by M_2 , M_3 and M_1 .

The four models specify person parameters as random-effects which follow a Normal distribution and assume one latent factor. The estimates and their standard errors for models with constant thresholds M_1 and M_2 are shown in Table 5. Both models agree in indicating exactly the same ranking of the items in terms of 'how easy it is to score higher categories' (measured throughout the *item parameter* β_i). The value of 'reading' has been set to 0 in order to estimate freely the others *threshold* and *item parameters* and to collocate them on a scale $(-\infty, +\infty)$. 'Reading' has been selected as reference item since it shows the most heterogeneous rate of responses across the three categories (0.34, 0.33, 0.33). From Tables 5, 6, 7 arise that 'attendance at lecture' is the easiest item, followed by 'no work activity' and 'clubbing'. The most difficult are the items at the bottom of the ranking: 'the use of university canteen', 'the use of sportive facilities' and 'attendance at theater'. The value of *item parameters* agree with the results arisen from the descriptive analysis where items 'attendance at theater' and 'attendance at lecture' respectively exhibit the lowest and the highest rate of responses in category *never*.

Fixing the load of 'reading' equal to 1 and leaving the others free to vary arises that the three aspects 'attendance at theater' (0.99), 'the use of university sport facilities' (0.89) and 'attendance at disco' (0.90) have a discrimination power close to 1, whereas items 'sport' (1.97), 'bar' (1.54) and 'clubbing' (1.37) discriminate more between subjects with different level of 'quality of life'. The lowest factor loadings are attached to aspects strictly linked to academic activities: 'use of university canteen' (-0.04), 'no work' (0.006), 'attendance at lecture' (0.126) and 'use of the university library' (0.25).

The analysis continues by comparing results of M_3 and M_4 . The estimated cut-points for both models are given in Table 5 and 6. The *proportional odds model* with different *threshold parameters* across the items makes the ranking of the aspects in terms of 'easiness' not unique since they can be sorted according to the values of the cut-points on the first or on the second category. To make easier the comparison of the rankings (according to their level of easiness obtained) sorted out using the four modeling approaches we evaluate the *Spearman Correlation Coefficient*.

Looking at the first cut-point a level of agreement equal to 0.95 is detected between the ranking provided by M_3 and M_4 , 0.94 between (M_1, M_2) and M_3 and 0.97 between (M_1, M_2) and M_4 . On the second cut point the ranking provided by M_3 still shows a good agreement ($\rho = 0.95$) with the results arisen under (M_1, M_2) . The agreement with M_4 is sensibly weaker ($\rho = 0.73$). The model to scale

Table 5: Graded Response Models: M_1 and M_2

items	M_1		M_2	
	τ_1	τ_2	τ_1	τ_2
	β_i	α_i	β_i	α_i
lectures	-0.977 (0.110)	1.061 (0.111)	-1.025 (0.123)	1.136 (0.123)
no work	1.123 (0.145)	1	1.140 (0.154)	0.126 (0.102)
bar	0.742 (0.144)	1	0.766 (0.157)	0.006 (0.104)
library	0.637 (0.141)	1	0.750 (0.154)	1.549 (0.253)
clubbing	0.484 (0.140)	1	0.499 (0.149)	0.251 (0.099)
reading	0.349 (0.140)	1	0.390 (0.149)	1.370 (0.231)
cinema	0.000 (0.000)	1	0.000 (0.000)	1.000 (0.000)
sport	-0.206 (0.138)	1	-0.196 (0.142)	0.635 (0.121)
disco	-0.435 (0.145)	1	-0.642 (0.177)	1.971 (0.327)
meeting	-0.839 (0.144)	1	-0.889 (0.150)	0.095 (0.171)
cultural	-1.177 (0.144)	1	-1.138 (0.151)	0.276 (0.102)
canteen	-1.475 (0.148)	1	-1.518 (0.156)	0.754 (0.155)
cus	-1.673 (0.154)	1	-1.617 (0.165)	-0.046 (0.110)
theater	-1.817 (0.156)	1	-1.915 (0.171)	0.889 (0.201)
	-2.366 (0.164)	1	-2.574 (0.198)	0.995 (0.206)

Table 6: Graded Response Model M_3

Items	thres. par.		discr. par.
	$\tau_{i(1)}$	$\tau_{i(2)}$	α_i
lectures	- 2.260 (0.175)	-0.020 (0.117)	1
library	-1.866 (0.155)	0.804 (0.123)	1
bar	-1.715 (0.147)	0.475 (0.118)	1
no work	-1.510 (0.142)	0.220 (0.118)	1
clubbing	-1.446 (0.138)	0.796 (0.122)	1
cinema	-1.211 (0.133)	1.840 (0.150)	1
reading	-0.742 (0.123)	0.834 (0.123)	1
sport	-0.205 (0.118)	1.019 (0.126)	1
disco	-0.119 (0.118)	1.861 (0.152)	1
meeting	0.058 (0.118)	2.919 (0.219)	1
cultural	0.491 (0.120)	2.595 (0.192)	1
canteen	0.803 (0.123)	2.162 (0.169)	1
cus	0.944 (0.126)	2.314 (0.175)	1
theater	1.377 (0.136)	3.702 (0.300)	1

Table 7: Graded Response Model M_4

Items	thres. par.		discr. par.
	$\tau_{i(1)}$	$\tau_{i(2)}$	α_i
bar	-7.466 (2.346)	2.135 (0.900)	12.870 (4.621)
clubbing	-3.658 (0.570)	2.139 (0.532)	7.187 (1.686)
lectures	-2.073 (0.164)	-0.026 (0.104)	0.144 (0.169)
library	-1.700 (0.144)	0.725 (0.112)	0.379 (0.182)
no work	-1.387 (0.129)	0.209 (0.104)	0.078 (0.161)
cinema	-1.257 (0.152)	1.878 (0.170)	1.442 (0.315)
reading	-0.708 (0.125)	0.778 (0.125)	1.000 (fixed)
sport	-0.210 (0.143)	1.078 (0.154)	1.689 (0.364)
disco	-0.133 (0.143)	2.010 (0.186)	1.703 (0.377)
meeting	0.036 (0.106)	2.707 (0.213)	0.386 (0.182)
cultural	0.461 (0.123)	2.509 (0.200)	1.043 (0.268)
canteen	0.694 (0.110)	1.996 (0.159)	-0.181 (0.180)
cus	0.860 (0.120)	2.159 (0.173)	0.633 (0.224)
theater	1.437 (0.167)	3.804 (0.329)	1.455 (0.357)

M_2). The agreement with M_4 is sensibly weaker ($\rho = 0.73$). The model to scale the set of item has been chosen also by considering the degree of uncertainty related to its estimates. The high values of the standard errors for several *items* and *discrimination parameters* which characterize M_4 signal that their values are poorly determined, making the estimates of these parameters strongly unreliable³. Since it is recommended to be aware against placing undue weight on small inequalities when standard errors are fairly large in relation to the difference in estimates (Barholomew et al., 2002), the model has not be considered in further analysis. We preferred to define the latent variable that scales the responses to the set of items using the parameters provided by M_2 which is the second best model among the GRM in terms of goodness of fit measures⁴.

The estimates of the fixed and random parameters of models M_2 are used in order to get the posterior estimates of the person parameters that will be used as indicators of students' position on the latent trait 'quality of life' in the scale $-\infty, +\infty$. Figure 1 shows the estimate of the person parameter for each student and its 95% confidence interval ordered according to the expected value from the

³ The same drawback has been detected for the factor loadings of the GPCM which showed unreliable standard errors.

⁴ The GPCM has not been considered as second choice since its estimate are strongly unreliable. The PCM shows AIC and BIC higher than M_2 .

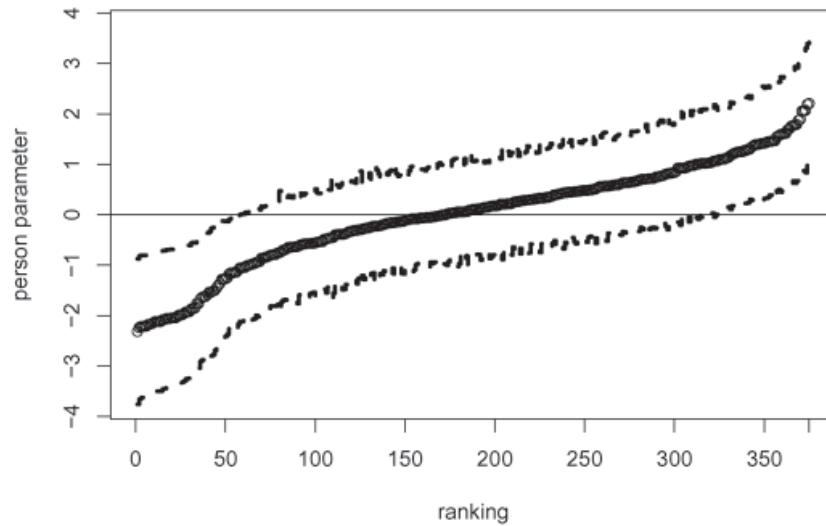


Figure 1: M_2 results: expected value of person parameters ‘student’s quality of life’ and pairwise 95% overlap intervals

eters of quality of university life if and only if their respective intervals do not overlap (Goldstein and Spiegelhalter, 1996).

It is interesting to stress the high values of the *Pearson Correlation Coefficient* between the expected values of person parameters estimated using M_1 ; M_2 ; M_3 models (all pairs of indexes show values greater than 0.90). Since the three models provide similar rankings of the students, we select M_2 which shows the best goodness of fit and higher variability of the expected values of the person parameters.

Figure 1 does not enable us to make a clear ranking across students but highlights the existence of clusters which are characterized by different levels of ‘quality of university life’. Three main groups can be detected: the first is composed by students who have the overall confidence interval below 0; the second highlights a fairly large cluster whose confidence intervals cross the 0; finally a third group shows confidence intervals which lie completely in the range of positive values.

4. A COMPARISON WITH OTHER SCALING METHODS FOR ORDERED VARIABLES

In the following the individual measures of students' quality of life obtained by using the specific modeling approach for categorical ordered data are compared with the values that we would observe by adopting two classical non scaling methods which have the main drawbacks to consider the category of the ordered variable equally spaced and/or to assign equal weights to all items which contribute to the definition of the synthetic indicator. In this attempt to make a comparison we are not considering the uncertainty on the final score associated to the mean value of the person parameters.

The first method assigns numbers in arithmetic progression to contiguous modalities, making implicitly the strong assumption of constant distance between adjacent categories. In our analysis each of the three categories of responses 'never', 'sometimes', 'often' has been replaced with numerical values '0.33, 0.66, 1.00'. The synthetic indicator for each individual has been obtaining using as merging function $M(w_1y_1, \dots, w_Iy_I)$ the un-weighted mean ($w_i = 1 \forall i$) of the I indicators. The second method, known as procedure of *undirect determined quantification* (Delvecchio, 2002), supposes that each ordinal item i is generated by a latent continuous variable z^* that we can not directly observed. The observed ordinal categories are linked to the latent z^* (Jöreskog, 2002; Torgerson, 1958) as follows

$$Y_{ip} = k \Leftrightarrow \gamma_{k-1} < z^* \leq \gamma_k, \quad k=1, \dots, K; \quad (4)$$

where Y_{ip} is the rating given to item i by unit p , $k = 1, \dots, K$ is the number of categories and γ_k are the cut points of the underlying continuous variable

$$-\infty \leq \gamma_0 \leq \gamma_1 \leq \gamma_2 \leq \dots \leq \gamma_{K-1} \leq \gamma_K \leq +\infty.$$

The distribution of the underlying latent variable z^* is supposed to be standard normal, with density function $\phi(z)$ and cumulative distribution function $\Phi(z)$. In this way, the underlying variable z^* assigns a metric to the ordinal categories. The percentage of responses in category k is given by

$$\pi_k = Pr[\gamma_{k-1} < z^* < \gamma_k] = \int_{\gamma_{k-1}}^{\gamma_k} \phi(u) du = \Phi(\gamma_k) - \Phi(\gamma_{k-1}). \quad (5)$$

The k -th category of the ordered item will be bounded by the quantiles γ_{k-1} and γ_k of the standard normal distribution that will be identified using the following relationships

$$\gamma_k = \Phi^{-1}(\gamma_k). \quad (6)$$

Thus the cumulative distribution of the percentage of responses in each of the K categories of item i is used to estimate the lower and upper bound of each category. The score y_k assigned to each ordinal category k corresponds to the median value of $\hat{\gamma}$ between the two extremes $\hat{\gamma}_{k-1}:\hat{\gamma}_k$ (Table 8).

The set of 14 items, scaled using estimates provided by Table 8, have been summarized in a single indicator (M_6) taking as merging function the un-weighted mean of the indicators.

Indicators M_2 , M_5 and M_6 have been made dimensionless by mapping their values in the range 0-1. This task has been pursued by adopting the re-scaling transformation function

$$m = \frac{M - \min^*(M)}{\max^*(M) - \min^*(M)}; \quad (7)$$

where $\max^*(M)$ and $\min^*(M)$ express the maximum and minimum expected values under the assumption that their values were equal to their maximum or to their minimum. For the standard Normal distribution the maximum and the minimum values have been set equal to the values -2.575 and 2.575 ($\Phi_{0.005}$ and $\Phi_{0.995}$).

The *Pearson Correlation Coefficient* between pairs of them (depicted in the upper panel of Figure 2) shows a high level of agreement between the indicators calculated by using the three different scaling methods (the value of ρ within pairs of indicators is always greater than 0.90). The greatest agreement is observed between m_5 and m_6 ($\rho = 0.97$) which assign equal weight to all indicators. However from an inspection of Figure 2 arises the existence of some clusters of units which assume difference position by using classical scaling methods M_5 and M_6 instead of the IRM approach M_2 .

The position indexes (Table 9) of the re-scaled indicators m_2 , m_5 and m_6 show that the first differentiates more across students with different quality of life since the discrimination parameter in model M_2 assigns a specific weight to each indicator involved in the analysis. Furthermore the standard deviation of the indicator m_2 is about twice the standard deviation of m_5 and m_6 since it discriminates more among students' with different response pattern. In an attempt to understand some interesting features that motivate the use of the IRM approach for ordered categorical variables, the pattern of responses of some units which show a remarkable different behavior adopting different scaling methods has been deeply

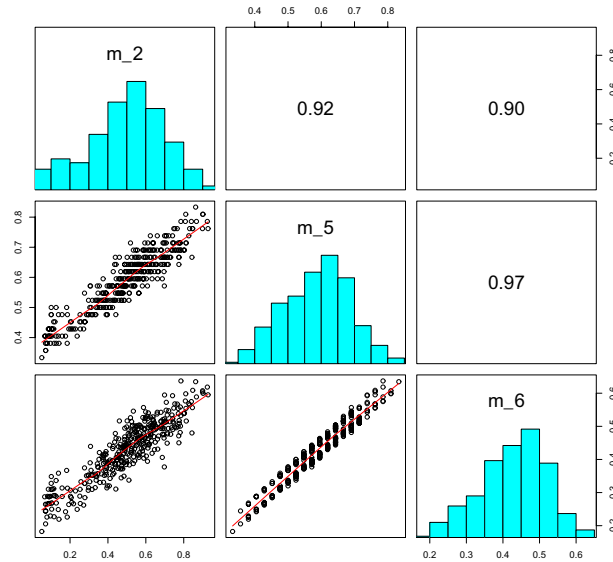


Figure 2: Scatter plot between pairs of synthetic indicators

Table 8: Estimates of threshold parameters

items	$\hat{\gamma}_{1.Med}$ never	$\hat{\gamma}_{2.Med}$ sometimes	$\hat{\gamma}_{3.Med}$ often
lectures	-1.589	-0.517	0.664
library	-1.414	-0.215	0.978
cus	-0.391	0.818	1.601
canteen	-0.431	0.750	1.555
meeting	-0.660	0.591	1.852
no work	-1.282	-0.316	0.759
sport	-0.750	0.198	1.045
cinema	-1.142	0.111	1.387
theater	-0.292	1.123	2.144
cultural events	-0.521	0.698	1.706
bar	-1.337	-0.291	0.823
disco	-0.724	0.398	1.405
reading	-0.957	0.007	0.967
clubbing	-1.223	-0.148	0.957

Table 9: Descriptive statistics for the three indicators

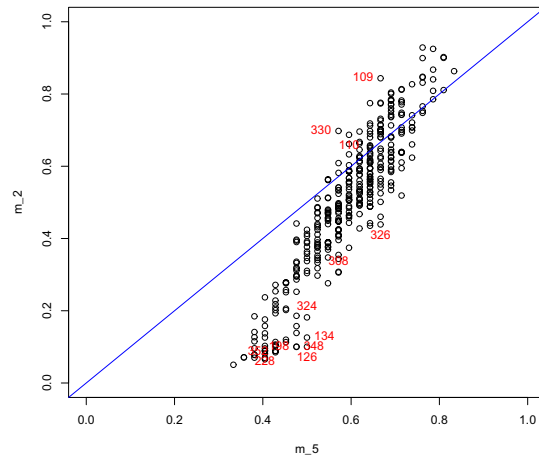
Position Index	M_2	M_5	M_6
Min.	-2.315	0.3333	-0.907
1 st Q	0.574	0.5476	-0.195
Median	0.108	0.6190	0.052
Mean	0.000	0.6135	0.015
3 rd Q	0.698	0.6905	0.266
Max.	2.208	0.8333	0.738
re-scaled indicators m_2, m_5, m_6			
Position Index	m_2	m_5	m_6
Min.	0.051	0.3333	0.182
1 st Q	0.389	0.5476	0.373
Median	0.521	0.6190	0.439
Mean	0.500	0.6135	0.428
3 rd Q	0.635	0.6905	0.496
Max.	0.928	0.8333	0.637
sd	0.198	0.100	0.089

investigated. In the following we will refer to them as ‘anomalous units’. Figure 3 highlights some of the units whose positions in the scale 0-1 are sensible to the choice of the scaling method.

The response patterns in Table 10 show that for units in cluster (a) the higher positions observed adopting m_5 instead m_2 are mainly determined by responses provided to items ‘no work’, ‘canteen’, ‘library’ and ‘lecture’. These items have an extremely low discrimination power (0.006, -0.046, 0.251, 0.126) and their overall influence in determining the final score of the synthetic indicator is marginal.

Other differences are detectable having a look at cluster (b) that shows response patterns of units which have the same value of the synthetic indicator m_5 . The difference between the response profiles of units 109 and 326 are strongly highlighted by m_2 which assumes a value equal to 0.843 for unit 109 and equal to 0.439 for unit 326. Subject 109 is involved in activities which have a greater discrimination power (‘sport’, ‘bar’, ‘clubbing’, ‘reading’, ‘theater’). This characteristic arises using the scale method M_2 which attaches higher levels of ‘quality of life’ to unit 109, whilst method M_5 does not enable us to discriminate across the items in terms of contribution to the overall measure of ‘quality of life’.

The last five units at the right end of Table 10 (c) show better positions on the scale of the synthetic indicator M_2 than M_5 . Units 330 and 110 provide most of the responses of the right end of the scale for items with particular high factor

Figure 3: Scatter plot between m_5 and m_2 Table 10: Response pattern of ‘anomalous’ units in respect of M_2 and M_5

items	a					b		c		
	b. p. ¹ under M_5					s. p. ² under M_5		b. p. under M_2		
	134	126	348	324	308	109	326	330	110	109
lectures	3	3	3	3	3	1	3	1	1	1
library	3	3	3	3	2	1	3	1	2	1
cus	1	1	1	1	2	2	3	1	1	2
canteen	2	3	3	3	3	1	3	1	1	1
meeting	2	1	1	1	2	1	3	1	1	1
no work	2	2	3	3	3	1	1	1	1	1
sport	1	1	1	1	1	3	1	3	3	3
cinema	1	1	1	1	3	3	2	1	1	3
theater	1	1	1	1	1	2	1	1	1	2
cultural	1	1	1	1	1	2	1	1	3	2
bar	1	1	1	1	1	3	2	3	3	3
disco	1	1	1	1	1	2	1	3	3	2
reading	1	1	1	2	2	3	2	3	1	3
clubbing	1	1	1	1	1	3	2	3	3	3

¹ b.p=better performance² s.p= same performance

loadings: 'sport' (1.971), 'club' (1.370), 'bar' (1.594), 'disco' (0.905). The value of unit 109 that is high in respect of 330 and 110 is determined by responses provided to items 'theater' and 'cinema'.

5. CONCLUSION

In this work a modeling approach for the assessment of 'students' quality of life' has been built up and tested on a sample of students' enrolled to three faculties of the University of Cagliari: Economics, Law and Political Sciences. The use of different scaling methods based on *Item Response Models* highlights the aspects which allow to differentiate more across subjects who have different habits of life. The method attaches the greatest discrimination power to all the activities not directly linked to the university life; specifically to attend frequently pubs, clubs, cultural events, theaters and sporting centers is what makes the difference between those students who are just involved in their academic studies and those who are perfectly integrated in the city, enjoy their students' status by taking part to external activities and try to take the greatest advantage from the city environment. An interesting point which has not been faced in this work concerns the association between the 'students' quality of life' and students' academic success; unfortunately, the unavailability of information did not allow us to perform further investigation on this aspect. Another issue left unexplored is the dimensionality of the set of indicator variables adopted to define the latent trait. The determination of the number of the factors underling the latent trait 'students' quality of life' is a delicate issue that we overcame by allowing the 14 indicators to load on one factor.

The main advantage of the IRM methodological approach is that the ordinal scale of the items is specifically taken into account in the estimation of threshold parameters and thus no arbitrary assumptions are advanced on the distance between adjacent categories. Nevertheless, the method indirectly attaches different weights (loads) to the items in the definition of the latent variable leaving the discrimination parameters free to vary. Comparisons between these methods and two 'classical' scaling methods adopted in the past to score ordered variables revealed that even though there is a high level of agreement between rankings sorted out using the three approaches, M_5 and M_6 poorly discriminate between subjects which are differently involved in the activities described by the bunch of the indicator items. The analysis carried out shows that IRM for polytomous ordered items turns out to be a useful research tool in the phase of definition of a suitable function to summarize information in a composite indicator.

References

- Agresti, A. (2002). *Categorical Data Analysis*. Wiley-Interscience, Hoboken, New York.
- Agresti, A., Booth, G., Hobert, O., and Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology*, (30): 27–80.
- Aureli, E. and Grimaccia, E. (1999). Un percorso metodologico per lo studio della qualità della vita universitaria degli studenti. In E. Aureli, F. Buratto, L. Carli Sardi, A. Franci, A. Ponti Sgargi, and S. Schifini D'Andrea, eds., *Contesti di Qualità della Vita. Problemi e Misure*, 122–154. Franco Angeli, Milano.
- Baker, F.B. and Kim, S.H. (2004). *Item Response Theory: Parameter Estimation Techniques*. Dekker, New York.
- Balbi, S. and Grassia, M.G. (2007). Profiling and labour market accessibility for the graduates in economics at Naples university. In L. Fabbri, ed., *Effectiveness of University Education in Italy*, 345–356. Physica-Verlag, Heidelberg.
- Bartholomew, D.J. (1998). Scaling unobservable constructs in social science. *Applied Statistics*, (47): 1–13.
- Bartholomew, D.J., Steele, F., Moustaki, I., and Galbraith, J.I. (2002). *The Analysis and Interpretation of Multivariate Analysis for Social Scientists*. Chapman & Hall, Boca Raton.
- Bernardi, L., Capursi, V., and Librizzi, L. (2004). Measurement awareness: The use of indicators between expectations and opportunities. In *Atti XLII Convegno della Società Italiana di Statistica. Bari, 9-11 Giugno 2004*, 315–326. Società Italiana di Statistica.
- Biggeri, A. and Bini, M. (2001). Evaluation at university and state level in Italy: Need for a system of evaluation and indicators. *Tertiary Education and Management*, (7): 149–162.
- Bini, M. and Bertaccini, B. (2007). Evaluating the university educational process. A robust approach to the drop-out problem. In L. Fabbri, ed., *Effectiveness of University Education in Italy*, 55–70. Physica-Verlag, Heidelberg.
- Capursi, V. and Porcu, M. (2001). La didattica universitaria valutata dagli studenti: un indicatore basato su misure di distanza fra distribuzioni di giudizi. In *Atti Convegno Intermedio della Società Italiana di Statistica 'Processi e Metodi Statistici di Valutazione', Roma 4-6 giugno 2001*, 17–20. Società Italiana di Statistica.
- Chiandotto, B. and Bacci, S. (2007). Measurement of university external effectiveness based on the use of acquired skills. In L. Fabbri, ed., *Effectiveness of University Education in Italy*, 89–104. Physica-Verlag, Heidelberg.
- Cox, D., Fitzpatrick, R., Fletcher, A.E., Gore, S.M., Spiegelhalter, D.J., and Jones, D.R. (1992). Quality of life assessment: can we keep it simple? *Journal of the Royal Statistical Society A*, (155): 353–393.
- De Boeck, P. and Wilson, M., eds. (2004). *Item Response Models: A Generalized Linear and Non Linear Approach*. Statistics for Social and Behavioral Sciences. Springer, New York.
- Delvecchio, F. (2002). *Statistica per la ricerca sociale*, chap. Dalla Qualità alla Quantità, 442–444. Cacucci Editore, Bari.
- Fayers, P.M. and Hand, D.J. (2002). Causal variables, indicator variables and measurement scales: An example from quality of life. *Journal of the Royal Statistical Society B*, (165): 233–261.
- Goldstein, H. and Spiegelhalter, D.J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society A*, (159): 385–443.

- Huebner, E., Valois, R., Paxton, R.J., and Drane, J. (2005). Middle school students' perceptions of quality of life. *Journal of Happiness Studies*, (6): 15–24.
- Jöreskog, K.G. (2002). *Structural Equation Modeling with Ordinal Variables Using LISREL*. <http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf>.
- Larsen, R., Diener, E., and Emmons, R.A. (1985). An evaluation of subjective well-being measures. *Social Indicators Research*, (17): 1–17.
- Maggino, F. and Schifini, S. (1999). Qualità della vita universitaria validazione di strumenti soggettivi. In E. Aureli, F. Buratto, L. Carli Sardi, A. Franci, A. Ponti Sgargi, and S. Schifini D'Andrea, eds., *Contesti di Qualità della Vita. Problemi e Misure*, 155:184. Franco Angeli, Milano.
- Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, (47): 149–174.
- Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, (14): 59–71.
- Muraki, E. (1992). A generalized partial credit model: an application of an EM algorithm. *Applied Psychological Measurement*, (16): 159–176.
- Porcu, M. and Puggioni, G. (2003). Laurea e abbandono universitario. Uno studio comparativo dei due eventi su un campione di immatricolati dell'Università di Cagliari. In L. Fabbri, ed., *LAID-OUT: scoprire i rischi con l'analisi di segmentazione*, 25–40. Cleup, Padova.
- Porcu, M. and Tedesco, N. (2007). Transition from university to the job market. A time analysis of university of Cagliari graduates. In L. Fabbri, ed., *Effectiveness of University Education in Italy*, 183–194. Physica-Verlag, Heidelberg.
- Rampichini, C., Grilli, L., and Petrucci, A. (2004). Analysis of university course evaluations: From descriptive measures to multilevel models. *Statistical Methods & Applications*, (13): 357–371.
- Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, (17): 1:100.
- Shulz, W. (1999). Explaining quality of life: The controversy between subjective and objective variables. *EuReporting Working Paper 10*, European Commission, Bruxelles.
- Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variables Modeling*. Chapman & Hall, Boca Raton.
- Sulis, I. (2007). *Measuring Students' Assessments of 'University Course Quality' using Mixed-Effects Models*. Ph.D. thesis, Università degli Studi di Palermo, Palermo.
- Torgerson, W.S. (1958). *Theory and Methods of Scaling*. Wiley and Son, New York.